Robust Sampling Time Design for Biochemical Systems

Hui Yu*, Hong Yue,* Peter Halling**

* *Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, UK (e-mail: hui.yu.100@strath.ac.uk; hong.yue@strath.ac.uk)*

** *WestCHEM Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow G1 1XL, UK (e-mail: p.j.halling@strath.ac.uk)*

Abstract: Optimal sampling time design by considering parameter uncertainties has rarely been considered in published research. In this work, the robust experimental design (RED) for sampling time selection is investigated. The aim is to exploit the sampling strategy using which the experiment can provide the most informative data for improving parameter estimation quality. With an enzyme reaction case study system, two global sensitivity analysis (GSA) approaches, the Morris screening method and the Sobol's method, are firstly applied to find out the key parameters that have large influences to model outputs of interest. Then three different RED methods, the worst-case strategy, the Bayesian design, and the GSA-based approach, are developed to design the optimal sampling time schedule. Simulation results suggest that, among the three RED methods, the equally spaced sampling from the Bayesian design has the best robustness towards parameter uncertainties.

Keywords: optimal experimental design (OED), robust experimental design (RED), sampling time design, global sensitivity analysis (GSA), parameter estimation, enzyme reaction system.

1. INTRODUCTION

Optimal experimental design (OED) is a useful technique for data-based modelling, especially for biochemical and biological systems with complex dynamics. Performing experiments for such systems is usually expensive and time-consuming, it is therefore helpful to design the experiments by taking into account requirements for parameter estimation so that the data collected will be of good value. In model development of a biochemical system, typical aspects in experimental design include the initial input conditions, the external time-varying input profiles, the sampling time schedule, the measurement set selection, etc. A number of OED methods have been reported in recent works (Gil et al., 2014; Hagen et al., 2013; Yu et al., 2015; Paquet-Durand et al., 2015), most of which lead to local optimum results that depend on the current estimates of parameter values. In practice, model parameters of complex systems are usually only roughly known in priori, the OED results can therefore be over-optimistic in certain cases. It is requisite to develop efficient methods of robust experimental design (RED) that can involve model uncertainties in the design stage.

Model-based OED is mostly developed on the Fisher Information Matrix (FIM), which is a measure of information contained in data. The FIM can be calculated from local parametric sensitivities, i.e., the first-order partial derivatives of the model outputs over the parameters. Parameter sensitivities are functions of model parameters which means that the FIM is also determined by parameter values. For this reason, experimental designs based on FIM properties are regarded as local designs. The effectiveness of the design depends on how close the assumed model parameters are to the real values. Furthermore, important parameters selected based on local sensitivity analysis (LSA) depend also on the assumed parameter values. When model parameters contain large uncertainties, it is necessary to analyse parameter effects, on model outputs, through the whole parameter range rather than at a local set. Therefore, global sensitivity analysis (GSA) is required to comprehensively investigate parameter effects on model outputs of interest.

Different from the popular OED development in the past decades, there are very few developments of RED. One big challenge in RED is how to formulate the RED problem including a quantitative representation of model uncertainties. Another challenge is that the computational load for solving RED problems is much higher than OED because the whole parameter range, rather than a certain parameter set, need to be examined in RED. Most RED methods are developed from extension of OED methods, for example, a robust measurement set design is developed for a biochemical pathway model (He et al., 2010; Brown et al., 2008). Other work on the robust design of input conditions can be found in (Vanlier et al., 2012; Mdluli et al., 2015; Telen et al., 2013). To the best of our knowledge there is no work reported on the robust sampling time design by considering parameter uncertainties. In this work, we aim to investigate RED for complex biochemical systems, with a particular focus on the design of sampling time profile for a case study enzyme reaction system. To support the RED with model uncertainties, GSA rather

than LSA will be employed to examine the parameter effects on model outputs.

The rest of this paper is organised as follows. In Section 2, preliminaries on general model description and OED are briefly introduced. Two GSA methods are discussed in Section 3 for sensitivity analysis of model parameters with large uncertainties. Three robust design strategies, the worst-case approach, the Bayesian design, and a GSA-based method, are investigated for the sampling time selection in Section 4. Then the RED of sampling time design for an enzyme reaction system and the analysis of simulation results are presented in Section 4. Conclusions and future work are discussed in Section 5.

2. PRELIMINARIES

2.1 Mathematical Model Representation

A continuous dynamic model written in ordinary differential equations is one of the most prominent model descriptions for chemical and biochemical systems. Consider a general biochemical model with n state variables and p parameters, the state transition and output models can be represented as follows:

$$\begin{aligned} \dot{\mathbf{X}}(t) &= \mathbf{f}(\mathbf{X}(t), \boldsymbol{\theta}), \ \mathbf{X}(t_0) = \mathbf{X}_0 \\ \mathbf{Y} &= \mathbf{h}(\mathbf{X}(t), \boldsymbol{\theta}) + \boldsymbol{\xi}(t) \end{aligned} \tag{1}$$

where $\mathbf{f}(\cdot)$ is a set of state transition functions of the system dynamics which are assumed to be continuous and first-order derivative; $\mathbf{X} = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$ denotes the vector of n state variables with initial condition $\mathbf{X}_0$; $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_p]^T \in \mathbb{R}^p$ is the vector of p model parameters; $\mathbf{Y} = [y_1, y_2, \ldots, y_m]^T \in \mathbb{R}^m$ is the measurement output vector with m ($m \leq n$) measurable variables, and $\mathbf{h}(\cdot)$ is the measurement function, normally used for selecting which variables to be measured. $\boldsymbol{\xi}$ is the vector of measurement errors which can be classified into systematic errors and random errors. The experiments should be designed to eliminate the systematic errors. However, the random errors that disturb the observations always exist. Most often the measurement error is assumed to be a zero mean, Gaussian white noise.

2.2 Basics of Optimal Experimental Design

The Fisher information matrix (FIM) is widely used in OED as the basis to quantify the information content of parameter estimation. When the design factors that characterise the experiment are denoted as $\boldsymbol{\phi}$, the FIM can be expressed as (Zullo, 1991):

$$\mathbf{FIM}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \mathbf{S}(\boldsymbol{\theta}, \boldsymbol{\phi})^T \mathbf{WS}(\boldsymbol{\theta}, \boldsymbol{\phi}) \tag{2}$$

where the weighting matrix $\mathbf{W}$ quantifies the reliability of measurement data at every specific point, which is normally taken as the inverse of the measurement error covariance matrix. $\mathbf{S} = \partial\mathbf{X}/\partial\boldsymbol{\theta}$ is the local parametric sensitivity matrix representing the local effect of parameters on model outputs. Based on the Cramer-Rao inequality, the FIM is approximately equal to the inverse of parameter estimation error covariance matrix, thus provides the local lower bound of the variance for parameter estimates. The OED problem can be expressed as the optimisation of a proper measure of parameter error covariance matrix, i.e.

$$\boldsymbol{\phi}^* = \arg\min_{\boldsymbol{\phi} \in \boldsymbol{\Phi}} \upsilon\left(\left(\mathbf{FIM}^{-1}(\boldsymbol{\theta}, \boldsymbol{\phi})\right)\right) \tag{3}$$

where $\boldsymbol{\Phi}$ is the admissible space of the design factors, $\upsilon(\cdot)$ represents a scalar function of the inverse of FIM. The most commonly used design criteria are A-optimal, D-optimal, E-optimal, and modified E-optimal design (Ljung, 1998).

3. GLOBAL SENSITIVITY ANALYSIS

Global sensitivity approaches are alternatives to LSA to quantify parameter effects on model outputs, which are especially useful when large parameter uncertainties are contained in the model. The main advantage of GSA over LSA from the experimental design point of view is that parameters can be varied simultaneously, rather than individually, over their entire uncertainty range to investigate their effects on the outputs. GSA aims at apportioning the output uncertainty to the uncertainty in the model parameter values. The effect on the outputs of changing one parameter while all the others are varied as well is evaluated by GSA and this can help in discovering parameter interactions in a model. In this work, the following two GSA methods are employed. For brevity we ignore time and state variables and the model is expressed as $f(\boldsymbol{\theta})$.

3.1 Morris Screening Method

The Morris screening method (Morris, 1991) is a measurement tool of global sensitivity which is based on the so called elementary effect (EE). A number of values of EEs for each parameter can be obtained through a predefined randomly selected sampling strategy. The distribution of EEs from the i-th parameter is denoted as F_i. The sensitivities are measured in two parts: μ_i, the mean of EEs is an estimate of the overall effect of the i-th parameter on model outputs; σ_i, the standard deviation of EEs is an evaluation of the ensemble of influence of the i-th parameter, which is attributed to the interactions with other parameters. These two measures will be used to identify which parameters should be considered as important to (selected) outputs.

Consider a general model with p parameters and an output y, $y = f(\theta_1, \theta_2, \cdots \theta_p)$. Each parameter has an uncertainty region which is scaled from 0 to 1; and it may take values from $\{0, 1/(p-1), 2/(p-1), \cdots, 1\}$. Then the EE of the i-th parameter is defined as

$$EE_i(\boldsymbol{\theta}) = \sqrt{\left(\frac{f(\theta_1, \cdots, \theta_i + \Delta, \cdots, \theta_p) - f(\boldsymbol{\theta})}{\Delta}\right)^2} \tag{4}$$

where Δ is a predetermined multiple of $1/(p-1)$ and is taken to be $\Delta = 2/(2p-2)$. Producing a value for F_i requires random selection of a value for each θ_i from the grid and evaluation of y twice, one at the selected parameter values, the other after increasing θ_i by Δ. The difference between these two runs yields one EE term. The calculation will be repeated γ times to produce a random sample of γ EEs for F_i. This method is computationally cheap as it requires a relatively small number of model evaluations compared to other GSA approaches.

3.2 Sobol's Method

Sobols method (Sobol, 2001) is a popular GSA technique based on analysis of variance. In general, variance-based sensitivity analysis methods aim to quantify the amount of variance that each parameter contributes to the total variance of model outputs. With Sobol's sensitivity analysis, the variances caused by either a single parameter or by the interactions of two or more parameters are expressed as sensitivity indices. For a model represented by (1) with p parameters, the model function can be decomposed into summation of increasing dimensionality:

$$\begin{aligned} f(\boldsymbol{\theta}) =& f_0 + \sum_{i=1}^{p} f_i(\theta_i) + \sum_{i=1}^{p}\sum_{j=i+1}^{p} f_{i,j}(\theta_i,\theta_j) + \cdots \\ &+ f_{1,2,\cdots,p}(\theta_1,\cdots,\theta_p) \end{aligned} \tag{5}$$

where f_0 equals to the expectation value of the output. The total variance can then be determined as:

$$V = \int_{\Omega^p} f^2(\boldsymbol{\theta}) d\boldsymbol{\theta} - f_0^2 \tag{6}$$

with Ω^p the p-dimensional hypercube space of model parameters. The partial variances, which are the components of the total variance decomposition, are computed from each of the terms in (5) as

$$V_{i_1,\cdots,i_k} = \int_{i_1} \cdots \int_{i_k} f_{i_1,\cdots,i_k}(\theta_{i_1},\cdots,\theta_{i_k}) d\theta_{i_1}\cdots d\theta_{i_k} \tag{7}$$

where $1 \le i_1 \le \cdots \le i_k \le p$. With the assumption that parameters are mutually orthogonal, the variance of outputs to parameters can be decomposed as:

$$V = \sum_{i=1}^{p} V_i + \sum_{i=1}^{p-1}\sum_{j=i+1}^{p} V_{i,j} + \cdots + V_{1,2,\cdots,p} \tag{8}$$

In this way, the variance contributions to the total output variance of individual parameters and parameter interactions can be determined. These contributions are characterized by the ratio of the partial variance to the total variance, the Sobol sensitivity indices, as follows:

First order SI: $S_i = \frac{V_i}{V}$

Second order SI: $S_{i,j} = \frac{V_{i,j}}{V}$

Total order SI: $S_{T_i} = S_i + \sum_{j\neq i} S_{i,j} + \cdots$

The first order index is a measure for the variance contribution of the individual parameter θ_i to the total model variance which is also called the main effect. S_{T_i} is the result of the main effect of θ_i and all its interactions with other parameters.

4. ROBUST SAMPLING TIME DESIGN

To handle model uncertainties in experimental design, the most often used approach is to design experiments in a sequential way by alternating parameter estimation and experimental design (Hering and Šimandl, 2010; Goujot et al., 2012). In each iteration, the knowledge of the system parameters is improved by estimation and this knowledge can then be used to improve the quality of the next OED. While this approach is intuitive, in practice, it could be difficult to perform many (sequential) experiments on the same system due to limited experimental resources. Also, it is not guaranteed that the parameters will converge to the true values after iterations.

Another approach to RED aims at determining the experiment that optimises the worst possible performance for any values of $\boldsymbol{\theta}$ belonging to the parameter domain $\boldsymbol{\Theta}$ (Körkel et al., 2004; Flaherty et al., 2006; Rojas et al., 2007). The RED optimisation problem can be formulated as

$$\boldsymbol{\phi}^* = \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\upsilon\left(FIM(\boldsymbol{\theta},\boldsymbol{\phi})\right)\right) \tag{9}$$

For a proposed design, find the model parameters for which the scalar value of the FIM is the smallest, i.e. the worst possible obtainable information content for this specific design is determined. Next, design the experiment which can maximize the scalar value of the FIM with the worst parameter combination. Using this technique, the prior information on the parameters is limited to the knowledge of the parameter domain, i.e. the upper and lower bounds of the parameters. No information on the probability distribution of parameters is necessary.

Instead of dealing with the parametric uncertainty with a worst-case maximin design strategy, an alternative way is to take account of parametric uncertainty by considering a prior distribution $p(\boldsymbol{\theta})$ of parameters in the design process, which leads to the Bayesian experimental design (BED) (Tulsyan et al., 2012; Murphy et al., 2003). BED is to quantify the statistical representation and treat the experimental design problem under the Bayesian framework. Different from classical OED for non-linear models which depends on the nominal parameter values, or the maximin design that is based on the worst-case parameters, BED is based on prior distribution of parameter estimates and their variance, rather than on the chosen single-point values. A BED thus incorporates more comprehensive knowledge of model parameter estimates. In the Bayesian design framework, an experiment is said to be optimal if it satisfies (Huan and Marzouk, 2014; Ryan et al., 2014):

$$\boldsymbol{\phi}^*_{ED} = \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \mathbb{E}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\upsilon\left(FIM(\boldsymbol{\theta},\boldsymbol{\phi})\right)\right) \tag{10}$$

where $\mathbb{E}$ represents the expected value of the scalar function υ over all possible parameter values. The expectation is calculated with respect to the prior distribution of parameters $p(\boldsymbol{\theta})$, over the entire parameter space $\boldsymbol{\Theta}$ and the maximisation is performed over the entire space $\boldsymbol{\Theta}$. Typically, numerical integration using a discretised version of the probability density function is used to evaluate the expectation. This solution is numerically rather burdensome if fine discretisation is used and many parameters are involved. Therefore, a multi-dimensional quadrature rule for approximating the multiple integral (over all parameters) can be applied. Another alternative would be to use Monte Carlo techniques to approximate the expectation, requiring however a large number of realisations before convergence can be achieved.

The connection between GSA and OED has been discussed recently in (Chu and Hahn, 2013), where the consistency condition for applying OED criteria to GSA results has been investigated. When a model is linear in model parameters, the design based on the global sensitivity matrix ($\mathbf{S}_G$) can be reduced to the conventional linear design based on the design matrix, that is,

$$\lim_{\Delta\boldsymbol{\theta}\to 0} \upsilon\left(\mathbf{S}_G^T\mathbf{S}_G\right) = \upsilon\left(\mathbf{S}_L^T\mathbf{S}_L\right) \tag{11}$$

One should note that not all GSA techniques satisfy the consistency condition. In this work, two different GSA methods are integrated to the OED procedure. The first one is given by the mean of the local sensitivities over the parameter uncertainty region, which can be expressed as

$$s_i = \mathbb{E}\left[\frac{\partial X}{\partial \theta_i}\right]. \tag{12}$$

The expectation of local sensitivities of θ_i over all its possible values represents the average effect of θ_i on X.

The second GSA method is based on the mean of squared sensitivity values which can avoid the cancellation of the negative effect values of parameter sensitivities, the formulation of which is given as

$$s_i = \sqrt{\left\{\frac{\mathbb{E}\left[\left(\frac{\partial X}{\partial \theta_i}\right)^2\right]}{var\left[\frac{\partial X}{\partial \theta_i}\right]} - 1\right\}} \tag{13}$$

5. CASE STUDY FOR AN ENZYME REACTION SYSTEM

5.1 Global Sensitivity Analysis

The effectiveness of the proposed RED method in Section 3 is examined with simulation to an an enzyme reaction system. The full model equations, the nominal values of the kinetic parameters and the initial condition of input variables can be found in (Yue et al., 2013). In this simulation study, the experimental length is set to be 6000 seconds, and a total number of 200 equally spaced sampling points along the whole reaction time are selected for GSA caculation. The Morris sensitivity method and the Sobols sensitivity method are applied in order to identify key parameters that are most influential on model outputs.

With the Morris screening sensitivity analysis, the lower and upper bounds for uncertainty ranges of model parameters are set to be 50% and 150% of their nominal values. The parameters are assumed to follow uniform distribution within their bounded ranges. The number of levels, p, and the repetition number of runs, γ, are set to be 6 and 100, respectively. The parameter effect on different model outputs by using the Morris screening method is shown in Fig. 1. The parameter ranking based on the mean values of EEs is listed in Table 1.

Table 1. Parameter importance ranking based on the mean value of elementary effects

Model outputs	Parameter ranking (descend)
S	$k_1,\ k_2,\ k_{-1},\ k_5W,\ k_{-2},\ k_4$
P	$k_1,\ k_2,\ k_{-1},\ k_5W,\ k_{-2},\ k_4$
N	$k_3,\ k_5W,\ k_{-3},\ k_4,\ k_{-4},\ k_1$
Q	$k_3,\ k_5W,\ k_{-3},\ k_4,\ k_{-4},\ k_1$
R	$k_5W,\ k_2,\ k_1,\ k_3,\ k_{-1},\ k_6$
All five outputs	$k_5W,\ k_2,\ k_1,\ k_3,\ k_{-1},\ k_{-5}$

It can be seen that for state variables S and P, parameters k_1, k_2 and k_{-1} are obviously the most important

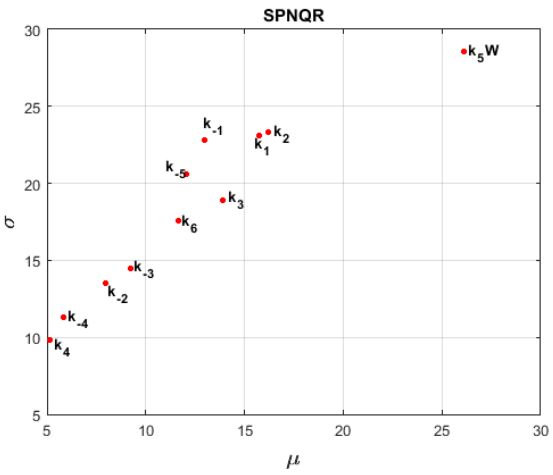

Fig. 1. Estimated mean and standard deviation of EEs for all 11 model parameters on all 5 model outputs

parameters. The effect of parameters on state variables N and Q are rather different, in which the most important parameters are found to be k_3 and k_5W. For state variable R, k_5W is found to be the more important parameter. By considering all these five model outputs, parameters $[k_1, k_2, k_5W]$ are identified to be the three most important parameters.

Next the Sobols GSA algorithm is applied to investigate the parameter effect based on the analysis of variance values. Simulation conditions on parameters and input variables are set to be the same as in Morris screening analysis. The Latin hypercube sampling strategy is employed where ten thousand samples for model parameters are selected for the analysis. The first order sensitivity measures and the total sensitivity measures by considering different model outputs are shown in Fig. 2. It can be observed that, for state variable S, k_2 is more important than other parameters; for state variable Q, the most important parameters are k_5W, k_{-3}, and k_3. This result is similar to the GSA from the Morris screening method.

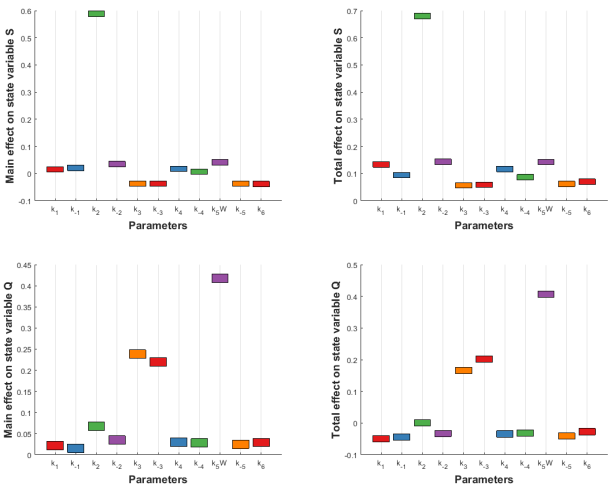

Fig. 2. Parameter effect to selected model outputs based on the analysis of variances, including the main effect and the total effect

5.2 Design of Sampling Time Schedule

Three different robust design methods are applied and compared for the sampling time design of the enzyme reaction system. Parameters k_2, k_{-3} and k_5W are selected in the design framework. In all three different experimental design simulations, the parameter uncertainty range is set to be 0.5-1.5 times of nominal parameter values. For the maximin robust design, the Powells method is applied for the optimal sampling time selection and the particle swarm algorithm is implemented to find the worst parameter values that will lead to the least informative data. In the Bayesian sampling time design, the Monte Carlo sampling strategy is employed to select ten thousand random parameter sets. In all three RED strategies, the D-optimal design criterion is used and the objective is to find twenty best sampling time points that can best facilitate parameter estimation when parameter uncertainties are considered. The RED results are listed in Table 2.

Table 2. Robust sampling time design with different RED techniques

RED methods	Sampling time points (sec)	D-values
Maximin design method	[330:30:510] [3300:30:3450] [4800:30:4980]	8.84e-8
Bayesian design method	[450:30:660] [2040:30:2100] [3060:30:3120] 3870 3900 4860 4890 5970 6000	6.02e-6
GSA based design (12)	[630:30:810] [5610:30:5970]	1.47e-5
GSA based design2 (13)	[1080:30:1260] [5610:30:5970]	0.0012

For the maximin robust design, the final obtained parameter values for the three selected parameters are $k_2 = 150$, $k_{-3} = 300$ and $k_5W = 7500$. By using this design method, the data information is maximised even if the worst parameter set is chosen. The exact sampling time points are selected at the start and middle stage of the reaction while the data information in the late stage of reaction is ignored (shown in Fig. 3). For the Bayesian robust sampling time design, the sampling time is loosely distributed along the whole reaction time and the averaged data information is maximised. The equally spaced sampling strategy is suggested. In other words, data information at each available sampling point has equal importance when model parameters contain large uncertainties. The GSA based experimental design can lead to more informative data. The exact sampling time points can be divided into two groups, in which the sampling time points should be selected at the start of the reaction and at the end of the reaction (shown in Fig. 3), in which each row represents one strategy (the scales in the vertical axis has no quantitative meaning)

In order to compare the design results from the 3 RED methods, we select 27 different parameter sets for k_2, k_{-3}, and k_5W, where each parameter can be chosen as 0.5, 1 and 1.5 times of their nominal values. The D-optimal values of different parameter sets with their corresponding optimal sampling strategies are listed in Table 3.

It can be seen from Table 3 that the results from the maxmin and the Bayesian designs are better compared to the two GSA-based designs. This is because the GSA is

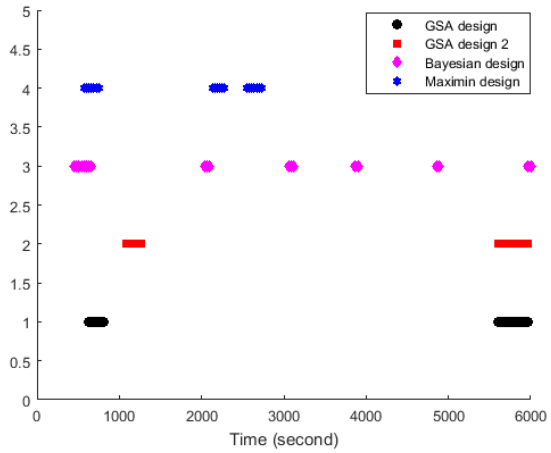

Fig. 3. Sampling time distribution obtained from three different RED methods

Table 3. Comparison of D-values among RED methods with various parameter sets

RED methods	best $ln(det(FIM))$	mean of $ln(det(FIM))$	standard deviations
Maximin design method	-34.16	-38.44	2.87
Bayesian design method	-33.80	-38.22	2.61
GSA based design (12)	-50	-52.49	2.07
GSA based design2 (13)	-33.75	-41.75	4.17

averaged along the whole range of parameter uncertainties, and there exists one certain parameter set, the sensitivities of which are close to the averaged GSA values. Therefore, the GSA based design result is close to the local design by using that particular parameter set. The performance of GSA based design is only good when parameter estimates are close to the true values. The standard deviation value also shows that this kind of design is sensitive to parameter values. The maximin design and Bayesian design, however, can provide higher D-values and smaller deviations, which indicate that these robust design methods are less sensitive to the change of parameter values and can provide in average higher data information than local design or GSA based designs. The Bayesian design is slightly better than the maximin design through the comparison of standard deviations, which implies that equally spaced sampling time selection is perhaps the best option when parameter uncertainties are large.

6. CONCLUSIONS

In this work, we have investigated RED for sampling time scheduling when model parameters are uncertain. Using a kinetically controlled synthesis process model, two GSA approaches, the Morris screening method and the Sobols sensitivity analysis, are implemented to determine crucial model parameters. The Morris method is based on the analysis of the EEs of model parameters on the outputs. The Sobol's method calculates the variance of the model

outputs attributed to parameter variations. Both GSA methods lead to similar results, in which k_2 and k_5W are found to be the most important parameters. Three RED methods are developed and applied to the sampling time design for an enzyme reaction system. Based on the simulation results, the maximin design can assure the maximal data information in the worst parameter set, while the Bayesian experimental design maximises the averaged data information along the whole range of parameter uncertainties. When GSA is integrated with the experimental design criteria, it can make the best performance in terms of the D-optimal value. However, by using this method, parameter uncertainties are averaged during sensitivity analysis. The final optimised sampling strategy might be associated to one particular parameter set (the one which is close to the averaged parameter sensitivity). This cancellation effect deteriorates the GSA-based RED performance. Through the comparison of the three RED simulation results, the equally spaced sampling time profile is demonstrated to be the best sampling strategy for systems with large parameter uncertainties.

Further investigations will be made to validate the robust sampling time design for other complex systems. Also, the integration of different design factors in one RED framework is a challenging task as the RED for different design factors are usually expressed as different optimisation problems. More efficient OED methods and numerical strategies need to be developed that could combine those design factors into one optimisation problem, so that the data information generated from designed experiment are less sensitive to model parameters.

ACKOWLEDGEMENT

This work is partially supported by Chinese NSFC project no. 61473025.

REFERENCES

Brown, M., He, F., and Yeung, L.F. (2008). Robust measurement selection for biochemical pathway experimental design. *Int. J. Bioinformatics Res. Appli.*, 4(4), 400–416.

Chu, Y. and Hahn, J. (2013). Necessary condition for applying experimental design criteria to global sensitivity analysis results. *Comput. Chem. Eng.*, 48(Suppl. C), 280–292.

Flaherty, P., Arkin, A., and Jordan, M.I. (2006). Robust design of biological experiments. In *Advances in Neural Information Processing Systems (NIPS)*, 363–370. MIT Press.

Gil, M.M., Miller, F.A., Silva, C.L., and Brandão, T.R. (2014). Application of optimal experimental design concept to improve the estimation of model parameters in microbial thermal inactivation kinetics. *J. Food Eng.*, 134, 59–66.

Goujot, D., Meyer, X., and Courtois, F. (2012). Identification of a rice drying model with an improved sequential optimal design of experiments. *J. Proc. Contr.*, 22(1), 95–107.

Hagen, D.R., White, J.K., and Tidor, B. (2013). Convergence in parameters and predictions using computational experimental design. *Interface Focus*, 3(4), 20130008.

He, F., Brown, M., and Yue, H. (2010). Maximin and bayesian robust experimental design for measurement set selection in modelling biochemical regulatory systems. *Int. J. Robust Nonl. Contr.*, 20(9), 1059–1078.

Hering, P. and Šimandl, M. (2010). Sequential optimal experiment design for neural networks using multiple linearization. *Neurocomputing*, 73(16), 3284–3290.

Huan, X. and Marzouk, Y. (2014). Gradient-based stochastic optimization methods in bayesian experimental design. *Int. J. Uncertain. Quantif.*, 4(6), 479–510.

Körkel, S., Kostina, E., Bock, H.G., and Schlöder, J.P. (2004). Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optim. Methods Softw.*, 19(3-4), 327–338.

Ljung, L. (1998). System identification. In *Signal Analysis and Prediction*, 163–173. Springer.

Mdluli, T., Buzzard, G.T., and Rundell, A.E. (2015). Efficient optimization of stimuli for model-based design of experiments to resolve dynamical uncertainty. *PLOS Computational Biology*, 11(9), 1–23.

Morris, M.D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2), 161–174.

Murphy, E.F., Gilmour, S.G., and Crabbe, M.J.C. (2003). Efficient and accurate experimental design for enzyme kinetics: Bayesian studies reveal a systematic approach. *J. Biochem. Biophys. Methods*, 55(2), 155–178.

Paquet-Durand, O., Zettel, V., and Hitzmann, B. (2015). Optimal experimental design for parameter estimation of the peleg model. *Chemom. Intell. Lab. Syst.*, 140, 36–42.

Rojas, C.R., Welsh, J.S., Goodwin, G.C., and Feuer, A. (2007). Robust optimal experiment design for system identification. *Automatica*, 43(6), 993–1008.

Ryan, E.G., Drovandi, C.C., Thompson, M.H., and Pettitt, A.N. (2014). Towards bayesian experimental design for nonlinear models that require a large number of sampling times. *Comput. Stat. Data Anal.*, 70(Suppl. C), 45–60.

Sobol, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Math. Comput. Simul.*, 55(1), 271–280.

Telen, D., Houska, B., Logist, F., Diehl, M., and Van Impe, J. (2013). Guaranteed robust optimal experiment design for nonlinear dynamic systems. In *Control Conference (ECC), 2013 European*, 2939–2944. IEEE.

Tulsyan, A., Forbes, J.F., and Huang, B. (2012). Designing priors for robust bayesian optimal experimental design. *J. Proc. Contr.*, 22(2), 450–462.

Vanlier, J., Tiemann, C.A., Hilbers, P.A.J., and van Riel, N.A.W. (2012). A bayesian approach to targeted experiment design. *Bioinformatics*, 28(8), 1136–1142.

Yu, H., Yue, H., and Halling, P. (2015). Optimal experimental design for an enzymatic biodiesel production system. *IFAC-PapersOnLine*, 48(8), 1258–1263.

Yue, H., Halling, P., and Yu, H. (2013). Model development and optimal experimental design of a kinetically controlled synthesis system. *IFAC Proceedings Volumes*, 46(31), 327–332.

Zullo, L. (1991). *Computer Aided Design of Experiments: An Engineering Approach*. Ph.D. thesis, Imperial College London.